













WORKING PAPER  
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

BAYES VS. THE LENS: A THEORETICAL CRITIQUE

by

Michael F. van Breda

WP 1034-78

December 1978

MASSACHUSETTS  
INSTITUTE OF TECHNOLOGY  
50 MEMORIAL DRIVE  
CAMBRIDGE, MASSACHUSETTS 02139





BAYES VS. THE LENS: A THEORETICAL CRITIQUE

by

Michael F. van Breda

WP 1034-78

December 1978



A question of interest to many is just how people make decisions. This involves issues such as how people select data, organize it, process it and what decision models they use. Studies have been made of topics such as the degree of self-insight people have into their decision process, the extent of functional fixation and the presence and nature of processing biases.

Two basic models have been explored in the literature.<sup>1</sup> The one derives from Bayes Theorem. The other relies on the Brunswick Lens model. Both are clearly sub-models of the more general decision theory approach.<sup>2</sup> The two have recently been contrasted by Wright (1977). This paper leans heavily on his work and seeks to extend and deepen his analysis.

Much of the empirical work involving these two models has relied on laboratory studies using students as subjects. The extent to which results derived in this manner can be generalized has been questioned in a general methodological sense by several writers. The analysis in this paper suggests that perhaps many of the empirical conclusions follow from subjects who have uniform priors and make utility free judgments. If so, the conclusions indeed might not be generalizable.

#### The Brunswick Lens Model

Brunswick<sup>3</sup> proposed his Lens Model as a method to relate environmental and individual-specific variables in the decision process. Typically the subject is provided with K realizations of N stimuli or equivalently a  $N \times K$  cue matrix  $\underline{X}$ .<sup>4</sup> Actual realizations of an observable variable  $y_e$  are then compared with the subject i's estimates, conditional on  $\underline{X}$ ,  $y_i$ . These are two  $N \times 1$  vectors. Inter alia the two sets of dependent variables are correlated creating the so-called achievement

index  $r_{y_e y_i}$ .

In general, we have  $y_e = f_e(X)$  and  $y_i = f_i(X)$ . However, it is usually assumed that these functions are linear so that

$$\hat{y}_e = X \hat{\beta}_e \quad ; \quad \hat{y}_e : N \times 1, \hat{\beta}_e : K \times 1 \quad (1)$$

$$\hat{y}_i = X \hat{\beta}_i \quad ; \quad y_i : N \times 1, \hat{\beta}_i : K \times 1 \quad (2)$$

The vectors of beta weights are assumed to indicate the relative importance of each of the stimuli for the subject and in the environment.

There is a question here whether the linear models so derived are in some sense actually used by the subject or whether they more properly constitute an ex post fit only. The question is probably irrelevant since what we are after is an ability to predict a subject's response and not, even were it possible, to trace the exact synaptic chain in the neurological system.

It is of interest to note, however, that the linearity assumption appears to work remarkably well in its description of human judgement in these empirical studies. Wright (1977), for example, reports R-squared's based on a linear regression of the subjective judgements against the stimuli as high as 95%. Clearly this is in part due to the actual environmental response being a linear function of the stimuli or cues. However, even where the actual function appears to be non-linear, subjects still appear to rely heavily on linear judgemental models. Why this should be so is patently a matter of interest.

Most of the studies involving the Lens model have, however, taken place in laboratory conditions. This is almost inevitable given the nature of the research. This does though raise the very real question as to whether these results can be generalized. Stated slightly differently, what implicit assumptions might be present in these laboratory

studies that might not be present in the real world? More generally, what is the relationship of this approach to the more general decision theory approach?

### Bayesian Probability:

We are on firmer theoretical ground here, albeit weaker empirical ground, than with the Lens Model. For good mathematical sense we require that

$$\Pr(y_n | \underline{x}_k) = \frac{\Pr(y_n) \Pr(\underline{x}_k | y_n)}{\sum_{j=1}^J \Pr(y_j) \Pr(\underline{x}_k | y_j)} \quad (3)$$

Or, in words, the probability of a given outcome conditioned on a set of stimuli  $\underline{x}_k$  is proportional to the unconditional probability of that outcome multiplied by the probability of the set of stimuli conditioned on that outcome.

The first of these probabilities, the unconditional probability, is most commonly known as the prior probability of the outcome. This is the probability assessment, in our context subjective, of course, of the occurrence of that particular outcome. The second of these probabilities is known as the likelihood function. It is the likelihood of that set of stimuli given the occurrence of that particular outcome.

Typically subjects are asked first to provide their priors on a set of outcomes and then to provide their posterior probabilities for the same set of outcomes based on the new information fed to them. Wright's survey reports that the results are mixed. For one, people do not appear to revise their priors sufficiently, i.e., their revisions are conservative when compared with the Bayesian rule.

### Decision theory:

The two paradigms as they stand are not really comparable. In essence, the one requires the subject to arrive at a best estimate of an

outcome  $y_n$  given a data vector  $\underline{x}_k$ . The other essentially requires the subject to arrive at an estimate of the probability of a given outcome  $y_n$  conditioned on the same data vector  $\underline{x}_k$ . Neither makes explicit mention of a loss or utility function.

A more complete analysis would involve a combination of probabilities and utility functions. In other words, decision theory would predict that in general people would select a given outcome based on the expected utility of that choice. By way of example, consider the following illustration drawn from Wonnacott and Wonnacott (1972). We are asked to estimate the length of a beetle given that our priors about the species are a mean length of 25 mm. with a variance of 4 mm., the distribution being assumed normal. Suppose now that a sample of 10 beetles yields an average of 20 mm. and a variance of 10 mm. A classical estimate of the 95% confidence interval would be

$$\theta = \bar{X} \pm 1.96 \frac{\sqrt{V}}{n}$$

$$= 20 \pm 1.96$$

However, it can be shown, and they do, that the posterior distribution is normal with

$$p(\theta|\bar{X}) = N \left\{ \frac{w_1 \bar{X} + w_2 \theta_0}{w_1 + w_2}, \frac{1}{w_1 + w_2} \right\}$$

where the zero subscript indicates a prior and

$$w_1 = \frac{1}{\sqrt{V}^2 n}$$

$$w_2 = \frac{1}{\sqrt{V}^2 \theta}$$

It follows that the Bayesian mean is

$$\bar{x} = \frac{1(20) + 1/4(25)}{1 + 1/4} = 21$$

$$\text{with a variance} = \frac{1}{1 + 1/4} = 0.8$$

To these results we now need to add a loss function. It is well known that a quadratic loss function implies a mean as an optimal estimate. The best estimate then, given a revision of priors and a quadratic loss function, is 21 as opposed to the earlier point estimate of 20.

What is also apparent from this analysis is the fact that as one's priors become more diffuse so the second weight  $W_2$  goes to zero. The posterior then reduces to the classical formulation

$$p(\theta|\bar{X}) = N(\bar{X}, \sigma^2/n) \quad \dots(7)$$

Then, given a quadratic loss function, the best estimate is the sample mean of 20.<sup>5</sup> To arrive at this requires, however, the double assumption of a suitable loss function and uniform or diffuse priors.

The Bayesian paradigm does not call for an estimate, merely the posterior distribution of all possible estimates. As such, it does not involve the use of loss functions, as indeed it should not, if the subject is Savage rational.<sup>6</sup>

The lens paradigm, on the other hand, does call for an estimate although it, in turn, makes no assumptions about either a loss function or probability distributions, whether prior or posterior. However, it is well known that OLS involves the minimization of the square of the errors. This implies a use of quadratic loss functions - or at least this is one way that one can interpret the resulting linear function.



### Maximum Likelihood Estimates:

There is an alternative approach to all of this involving maximum likelihood estimators which might shed light on the empirical results obtained to date in the Lens studies. It also enables one to link the Bayesian approach more closely to the Lens approach.

Consider first an individual who is asked to come up with a best judgement of some parameter or dependent variable  $\theta$ . Assume further that he or she is provided with a data vector  $\underline{X}$ . Finally, assume, most importantly that the subject makes a utility free judgement. This is essentially what subjects are required to do in the Bayesian paradigm. Alternatively stated, they are basically asked to choose that  $\theta$  which maximizes

$$\Pr(\theta | \underline{X}, \underline{H}) = \Pr(\theta | \underline{H}) \Pr(\underline{X} | \theta, \underline{H}) / \Pr(\underline{X} | \underline{H}) \quad \dots (8)$$

where  $\underline{H}$  is the prior information available to the subject. But, since the denominator of (8) is a constant this reduces to

$$\max_{\theta} \Pr(\theta | \underline{H}) \Pr(\underline{X} | \theta, \underline{H}) \quad \dots (9)$$

But this is precisely the maximum likelihood estimator.

Thus, the first result that we arrive at is that if the subject's priors are vague, the Bayesian decision rule collapses into a Maximum Likelihood Estimate. Whether people actually do use MLE's is, of course, a matter for empirical investigation. However, it is highly probable that most students, who form the bulk of the laboratory population, do indeed have very vague priors about the items they are asked to estimate. Moreover, in the absence of any real incentive or punishment, it is not at all unlikely that they generate fairly utility free judgements. Given these two probabilities, it would not be surprising to find them seeking for an estimate or judgement which is in some sense "most likely" - that value most consistent with the data. In other words, it would not be



entirely improbable to find that laboratory subjects were using Maximum Likelihood estimates.

This surmise is further strengthened if we make the further assumption that the data vector  $\underline{X}$  is normally distributed, or at least, approximately normal in its distribution. It is well known that in this case the ML estimate is identical to the OLS estimate and that both are linear. This might be true for the data itself or for the subject's opinion of the data or both.<sup>7</sup> All that we require is for the subjective perception of the data to be reasonably normal to arrive at a linear judgemental rule.

#### Conclusion:

What we have arrived at then is the surmise that if certain assumptions hold, a linear judgemental model is wholly to be expected. These assumptions are 1) that a utility free judgement be made 2) that the subject have fairly uniform priors 3) that the data be perceived to be normally distributed and 4) that a maximum likelihood estimator is being used.

However, in the real world of harsh risks and pleasant rewards, it is highly unlikely that any of these assumptions will be met. Judgements are rarely utility free. Subjects have considerable priors in their areas of competence. Some would even label them fixations. The data might or might not be normally distributed. However, even if it were, in the absence of quadratic loss functions, a linear decision rule would be improbable.

The question then arises can we trace at a slightly deeper level what might be going on. Can we design experiments to check whether the above assumptions are indeed being met in the laboratory? Can we perhaps enrich the environments by more information and/or rewards and punishments

to test what happens when they are not met? And can we observe, more closely, some actual decisions to test whether the linear decision rule is merely a laboratory animal or more widely used? And, if so, why?

Footnotes

- 1) A full survey of the literature together with a bibliography may be found in Wright (1977).
- 2) A good discussion of decision theory especially as it relates to accounting may be found in Demski (1972).
- 3) See Brunswick (1952, 1956). Also Castellán (1972, 1973), Dudycha and Naylor (1966), Hammond, Hirsch, and Todd (1964), Hirsch, Hammond and Hirsch (1964), Stewart (1976) and Tucker (1964).
- 4) A letter with a bar underneath it indicates a vector or a matrix depending on the context.
- 5) An absolute loss function suggests the median as an optimal estimate while a (0, 1) loss function suggests a mode. Where the distribution is symmetric, these will, of course, coincide with the mean.
- 6) For more details, see Savage (1954).
- 7) A "true" Bayesian will not admit of the existence of distribution other than a subjective one so that this formulation would be redundant.



## References

- Ashton, R. H., "The Predictive Ability Criterion and User Prediction Models," The Accounting Review (October, 1974), 719-732 (1974b).
- Brunswick, E., The Conceptual Framework of Psychology, University of Chicago Press, 1952.
- \_\_\_\_\_, Perception and the Representative Design of Experiments, University of California Press, 1956.
- Castellan, N. J., Jr., "The Analysis of Multiple Criteria in Multiple-Cue Judgment Tasks," Organizational Behavior and Human Performance, 8 (1972), 242-261 (1972).
- \_\_\_\_\_, "Comments on the 'Lens Model' Equation and the Analysis of Multiple-Cue Judgment Tasks," Psychometrika, 38-1 (March, 1973), 87-100.
- Demski, J., Information Analysis. Addison-Wesley (1972).
- Dudycha, A. L. and Naylor, J. C., "The Effects of Variations in the Cue R Matrix upon the Obtained Policy Equation of Judges," Educational and Psychological Measurement, 26 (1966), 583-603.
- Edwards, W., "Dynamic Decision Theory and Probabilistic Information Processing," Human Factors, (April, 1962), 59-73.
- \_\_\_\_\_, "Conservatism in Human Information Processing" in B. Kleinmuntz (ed.), Formal Representation of Human Judgement, New York: Wiley, 1968.
- Hammond, K. R., Hursch, C. J., and Todd, F. S., "Analyzing the Components of Clinical Inference," Psychological Review, 71 (1964), 438-456.
- Hursch, C., Hammond, K. R., and Hursch, J. L., "Some Methodological Considerations in Multiple Cue Probability Studies," Psychological Review, 71 (1964), 42-60.
- Savage, L. J., The Foundations of Statistics, John Wiley and Sons, Inc., 1954.
- Slovic, P., and Lichtenstein, S., "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," Organizational Behavior and Human Performance (1971), 649-744.
- Stewart, T. R., "Components of Correlation and Extensions of the Lens Model Equation," Psychometrika, 41 (March, 1976), 101-120.
- Tucker, L. R., "A Suggested Alternative Formulation in the Developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd," Psychological Review, 71 (1964), 528-530.
- Wonnacott, T.H. & Wonnacott, R.J., Introductory Statistics for Business and Economics. John Wiley (1972).

Wright, W. F., "Cognitive Information Processing Biases: Implications for Producers and Users of Financial Information," unpublished manuscript, Graduate School of Business, Stanford University, August, 1976 (1976a).

\_\_\_\_\_, "Self-Insight into the Cognitive Processing of Financial Information," Research Paper No. 241 (Revised), Graduate School of Business, Stanford University, October, 1976 (1976b).

\_\_\_\_\_, "An Empirical Study of Financial Judgment Models," Research Paper No. 274 (Revised), Graduate School of Business, Stanford University, October, 1976 (1976c).

\_\_\_\_\_, "Financial Information Processing Models: An Empirical Study," Research Paper No. 246 (Revised), Graduate School of Business, Stanford University, December, 1976 (1976d).

\_\_\_\_\_, "A Comparative Analysis of the Lens and Bayesian Information Processing Paradigms," unpublished manuscript, Graduate School of Business, Stanford University, February, 1977.









JY 02 '88

OG 30 '88

OG 28 '89

MAR 31 1990

SEP 27 1991

ACME  
BOOKBINDING CO., INC.

SEP 6 1983

100 CAMBRIDGE STREET  
CHARLESTOWN, MASS.

MIT LIBRARY  
3 9080 004 524 481

MIT LIBRARY  
3 9080 004 524 507

1031-78

HD28.M414 no.1032- 78  
Kalwani, Manoh/Structure of repeat buy  
736263 D\*BKS 00069315

3 9080 001 240 651

HD28.M414 no.1032- 78 1979  
Kalwani, Manoh/Structure of repeat buy  
737571 D\*BKS 00137169

3 9080 002 050 018

HD28.M414 no.1033- 78  
Baldwin, Carli/Liquidity preference un  
736011 D\*BKS 00066623

3 9080 001 209 383

HD28.M414 no.1034- 78  
Van Breda, Mic/Bayes vs. the lens :  
736013 D\*BKS 00066626

3 9080 001 209 433

HD28.M414 no.1035- 78  
Simon, Hermann/Dynamics of price elast  
736015 D\*BKS 00066627

3 9080 001 209 458

HD28.M414 no.1036- 78  
Schmalensee, R/A simple model of risk  
736017 D\*BKS 00066625

3 9080 001 209 417

HD28.M414 no.1038- 79  
Choffray, Jean/Methodology for segment  
736019 D\*BKS 00066621

3 9080 001 209 342

HD28.M414 no.1039- 79  
Meldman, Jeffr/Decision support system  
827646 D\*BKS 00069325

3 9080 001 240 875

HD28.M414 no.1040- 79  
Kaplan, Nancy /A computer code for sol  
736021 D\*BKS 00066622

3 9080 001 209 367

